DETERMINING THE SIGNAL DIMENSION IN SECOND ORDER SOURCE SEPARATION

Joni Virta[1] and Klaus Nordhausen[2]

[1] *Aalto University School of Science, Finland*

[2] *Vienna University of Technology, Austria*

Supplementary Material

S1. Proofs

Proof of Lemma 1. By Assumption 3 we have,

$$\sqrt{T}(\hat{\mathbf{S}}_\tau - \mathbf{D}_\tau) = \mathcal{O}_p(1), \quad \text{for all } \tau \in \mathcal{T} \cup \{0\},$$

where $\mathbf{D}_0 = \mathbf{I}_p$. This instantly implies the equivalent result for the symmetrized autocovariance matrices $\hat{\mathbf{R}}_\tau$,

$$\sqrt{T}(\hat{\mathbf{R}}_\tau - \mathbf{D}_\tau) = \mathcal{O}_p(1), \quad \text{for all } \tau \in \mathcal{T}.$$

Let vect be the row-vectorization operator that takes the row vectors of a matrix and stacks them into a long column. That is, $\text{vect}(\mathbf{X}) \in \mathbb{R}^{mn}$ for any $\mathbf{X} \in \mathbb{R}^{m \times n}$ and $\text{vect}(\mathbf{AXB}) = (\mathbf{A} \otimes \mathbf{B})\text{vect}(\mathbf{X})$ for any $\mathbf{A} \in \mathbb{R}^{s \times m}, \mathbf{X} \in$

$\mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{n\times t}$. By linearizing and row-vectorizing the definition $\mathbf{0} = \sqrt{T}(\hat{\mathbf{S}}_0^{-1/2}\hat{\mathbf{S}}_0\hat{\mathbf{S}}_0^{-1/2} - \mathbf{I}_p)$ and using Slutsky's theorem, we obtain,

$$\hat{\mathbf{B}}\sqrt{T}\mathrm{vect}(\hat{\mathbf{S}}_0^{-1/2} - \mathbf{I}_p) = -\sqrt{T}\mathrm{vect}(\hat{\mathbf{S}}_0 - \mathbf{I}_p) + o_p(1),$$

where $\hat{\mathbf{B}} = \mathbf{I}_p \otimes \hat{\mathbf{S}}_0\hat{\mathbf{S}}_0^{-1/2} + \mathbf{I}_p \otimes \mathbf{I}_p$. As $\hat{\mathbf{B}} \to_p 2\mathbf{I}_{p^2}$, its inverse is asymptotically well-defined, allowing us to multiply the relation from the left by $\hat{\mathbf{B}}^{-1}$, after which Slutsky's theorem and Assumption 3 yield that $\sqrt{T}(\hat{\mathbf{S}}_0^{-1/2} - \mathbf{I}_p) = \mathcal{O}_p(1)$.

Linearize next as,

$$\begin{aligned}\sqrt{T}(\hat{\mathbf{H}}_\tau - \mathbf{D}_\tau) &= \sqrt{T}(\hat{\mathbf{S}}_0^{-1/2} - \mathbf{I}_p)\hat{\mathbf{R}}_\tau\hat{\mathbf{S}}_0^{-1/2} + \sqrt{T}(\hat{\mathbf{R}}_\tau - \mathbf{D}_\tau)\hat{\mathbf{S}}_0^{-1/2} \\ &\quad + \mathbf{D}_\tau\sqrt{T}(\hat{\mathbf{S}}_0^{-1/2} - \mathbf{I}_p),\end{aligned} \tag{S1.1}$$

where the right-hand side is by the previous convergence results expressible as $\sqrt{T}(\hat{\mathbf{R}}_\tau - \mathbf{D}_\tau) + \mathcal{O}_p(1)$. The first claim now follows after the division by $\sqrt{T}$ and the addition of $\mathbf{D}_\tau$ on both sides.

For the second claim we observe only the lower right $p_0 \times p_0$ corner block $\hat{\mathbf{H}}_{\tau 00}$ and write,

$$\hat{\mathbf{H}}_{\tau 00} = \hat{\mathbf{T}}_1^\top\hat{\mathbf{R}}_{\tau,-0}\hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_1^\top\hat{\mathbf{R}}_{\tau,\mathrm{off}}\hat{\mathbf{T}}_2 + \hat{\mathbf{T}}_2^\top\hat{\mathbf{R}}_{\tau,\mathrm{off}}^\top\hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_2^\top\hat{\mathbf{R}}_{\tau 00}\hat{\mathbf{T}}_2, \tag{S1.2}$$

where $(\hat{\mathbf{T}}_1^\top, \hat{\mathbf{T}}_2^\top)^\top$, $\hat{\mathbf{T}}_1 \in \mathbb{R}^{(p-p_0)\times p_0}$, $\hat{\mathbf{T}}_2 \in \mathbb{R}^{p_0\times p_0}$ denotes the final $p \times p_0$

column block of $\hat{\mathbf{S}}_0^{-1/2}$ and $\hat{\mathbf{R}}_\tau$ has been partitioned correspondingly as

$$\hat{\mathbf{R}}_\tau = \begin{pmatrix} \hat{\mathbf{R}}_{\tau,-0} & \hat{\mathbf{R}}_{\tau,\text{off}} \\ \hat{\mathbf{R}}_{\tau,\text{off}}^\top & \hat{\mathbf{R}}_{\tau 00} \end{pmatrix}.$$

These matrices satisfy $\hat{\mathbf{T}}_1 = \mathcal{O}_p(1/\sqrt{T})$, $\hat{\mathbf{T}}_2 - \mathbf{I}_{p_0} = \mathcal{O}_p(1/\sqrt{T})$, $\hat{\mathbf{R}}_{\tau,\text{off}} = \mathcal{O}_p(1/\sqrt{T})$ and $\hat{\mathbf{R}}_{\tau 00} = \mathcal{O}_p(1/\sqrt{T})$ and we can write (S1.2) as

$$\begin{aligned}\hat{\mathbf{H}}_{\tau 00} &= \hat{\mathbf{T}}_1^\top \hat{\mathbf{R}}_{\tau,-0} \hat{\mathbf{T}}_1 + \hat{\mathbf{T}}_1^\top \hat{\mathbf{R}}_{\tau,\text{off}} (\hat{\mathbf{T}}_2 - \mathbf{I}_{p_0}) + \hat{\mathbf{T}}_1^\top \hat{\mathbf{R}}_{\tau,\text{off}} \\ &+ (\hat{\mathbf{T}}_2 - \mathbf{I}_{p_0})^\top \hat{\mathbf{R}}_{\tau,\text{off}}^\top \hat{\mathbf{T}}_1 + \hat{\mathbf{R}}_{\tau,\text{off}}^\top \hat{\mathbf{T}}_1 + (\hat{\mathbf{T}}_2 - \mathbf{I}_{p_0})^\top \hat{\mathbf{R}}_{\tau 00} (\hat{\mathbf{T}}_2 - \mathbf{I}_{p_0}) \\ &+ \hat{\mathbf{R}}_{\tau 00} (\hat{\mathbf{T}}_2 - \mathbf{I}_{p_0}) + (\hat{\mathbf{T}}_2 - \mathbf{I}_{p_0})^\top \hat{\mathbf{R}}_{\tau 00} + \hat{\mathbf{R}}_{\tau 00} \\ &= \hat{\mathbf{R}}_{\tau 00} + \mathcal{O}_p(1/T),\end{aligned}$$

concluding the proof. □

Proof of Lemma 2. The SOBI-solution is found as $\hat{\mathbf{U}}^\top \hat{\mathbf{S}}_0^{-1/2}$ where the orthogonal matrix $\hat{\mathbf{U}}$ is the maximizer of

$$\hat{g}(\mathbf{U}) = \sum_{\tau \in \mathcal{T}} \left\| \text{diag}\left(\mathbf{U}^\top \hat{\mathbf{H}}_\tau \mathbf{U} \right) \right\|^2. \tag{S1.3}$$

Let $\hat{\mathbf{U}}$ be a sequence of maximizers of (S1.3) and partition $\hat{\mathbf{U}}$ in the blocks $\hat{\mathbf{U}}_{ij}$ in a similar way as in the problem statement (ignoring the sequence of permutations $\hat{\mathbf{P}}$ for now). The proof of the lemma is divided into two parts. First, we establish the consistency of the off-diagonal blocks, $\hat{\mathbf{U}}_{ij} \to_p \mathbf{0}$, and, second, we show the rate of convergence, $\sqrt{T}\hat{\mathbf{U}}_{ij} = \mathcal{O}_p(1)$. That the diagonal blocks $\hat{\mathbf{U}}_{ii}$ are stochastically bounded follows simply from the compactness of the space of orthogonal matrices.

1. Consistency

Our aim to is to use a technique similar to the M-estimator consistency argument (Van der Vaart, 1998, Theorem 5.7), for which we need the Fisher consistency of the off-diagonal blocks, along with the uniform convergence of the sample objective function to the population objective function with respect to $\mathbf{U}$. By Fisher consistency we mean that all maximizers $\mathbf{U}$ of the population objective function,

$$g(\mathbf{U}) = \sum_{\tau \in \mathcal{T}} \left\| \mathrm{diag}\left(\mathbf{U}^\top \mathbf{H}_\tau \mathbf{U} \right) \right\|^2,$$

where $\mathbf{H}_\tau = \mathbf{S}_0^{-1/2} \mathbf{R}_\tau \mathbf{S}_0^{-1/2}$, can have their columns ordered to satisfy $\mathbf{U}_{ij} = \mathbf{0}$ for all $i \neq j$.

Recall the partitioning of the latent components in Section 3 into groups of sizes $p_1, \ldots, p_v, p_0$ and denote the autocovariance of the jth group for lag τ by $\lambda_{\tau j}$. The population autocovariance matrices satisfy,

$$\mathbf{S}_0 = \mathbf{I}_p \quad \text{and} \quad \mathbf{H}_\tau = \mathbf{R}_\tau = \mathbf{S}_\tau = \mathbf{D}_\tau,$$

where $\mathbf{D}_\tau = \mathrm{diag}(\lambda_{\tau 1} \mathbf{I}_{p_1}, \ldots, \lambda_{\tau v} \mathbf{I}_{p_v}, \mathbf{0}) \in \mathbb{R}^{p \times p}$ are diagonal matrices, $\tau \in$

$\mathcal{T}$. The population objective function has the upper bound,

$$g(\mathbf{U}) = \sum_{\tau \in \mathcal{T}} \left\| \mathrm{diag}\left(\mathbf{U}^\top \mathbf{D}_\tau \mathbf{U}\right) \right\|^2 \leq \sum_{\tau \in \mathcal{T}} \left\| \mathbf{U}^\top \mathbf{D}_\tau \mathbf{U} \right\|^2 = \sum_{\tau \in \mathcal{T}} \sum_{j=1}^{v} \lambda_{\tau j}^2 p_j, \quad \text{(S1.4)}$$

with equality if and only if $\mathbf{U}^\top \mathbf{D}_\tau \mathbf{U}$, $\tau \in \mathcal{T}$ are diagonal matrices, i.e. $\mathbf{U}$ is an eigenvector matrix of all $\mathbf{D}_\tau$, $\tau \in \mathcal{T}$. One such matrix is $\mathbf{U} = \mathbf{I}_p$ and the maximal value of $g(\mathbf{U})$ is thus indeed $\sum_{\tau \in \mathcal{T}} \sum_{j=1}^{v} \lambda_{\tau j}^2 p_j$.

We next show that a both sufficient and necessary condition for $\mathbf{W}$ to be a maximizer of g is that $\mathbf{W}$ has, up to the ordering of its columns, the form

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \mathbf{W}_{vv} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{W}_{00} \end{pmatrix}, \quad \text{(S1.5)}$$

where the partition into blocks is as in the statement of the lemma and the diagonal blocks $\mathbf{W}_{00}, \mathbf{W}_{11}, \ldots, \mathbf{W}_{vv}$ are orthogonal.

We start with a the “necessary”-part. Let $\mathbf{W}$ be an arbitrary maximizer of g and take its first column $\mathbf{w} = (\mathbf{w}_1^\top, \ldots, \mathbf{w}_v^\top, \mathbf{w}_0^\top)^\top$, partitioned in subvectors of lengths $p_1, \ldots, p_v, p_0$. As equality is reached in the inequality (S1.4) only when $\mathbf{U}$ is an eigenvector of all $\mathbf{D}_\tau$, we have that $\mathbf{D}_\tau \mathbf{w} = \pi_\tau \mathbf{w}$

for some $\pi_\tau \in \mathbb{R}$ for all $\tau \in \mathcal{T}$. It then holds for all τ that,

$$\mathbf{0} = (\mathbf{D}_\tau - \pi_\tau \mathbf{I}_p)\mathbf{w} = \begin{pmatrix} (\lambda_{\tau 1} - \pi_\tau)\mathbf{I}_{p_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & (\lambda_{\tau v} - \pi_\tau)\mathbf{I}_{p_v} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & -\pi_\tau \mathbf{I}_{p_0} \end{pmatrix} \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_v \\ \mathbf{w}_0 \end{pmatrix},$$

which yields the equation group,

$$\mathbf{0} = \begin{pmatrix} (\lambda_{\tau 1} - \pi_\tau)\mathbf{w}_1 \\ \vdots \\ (\lambda_{\tau v} - \pi_\tau)\mathbf{w}_v \\ -\pi_\tau \mathbf{w}_0 \end{pmatrix}.$$

We next proceed by proof through contradiction. Assume that two distinct subvectors of $\mathbf{w}$, say $\mathbf{w}_k$ and $\mathbf{w}_\ell$, both contain a non-zero element. Then

$$\lambda_{\tau k} = \pi_\tau \quad \text{and} \quad \lambda_{\tau \ell} = \pi_\tau, \quad \forall \tau \in \mathcal{T}, \tag{S1.6}$$

and we recall that $\lambda_{\tau 0} = 0$ for all $\tau \in \mathcal{T}$. The identities (S1.6) imply that $\lambda_{\tau k} = \lambda_{\tau \ell}$ for all $\tau \in \mathcal{T}$, i.e., that the kth and ℓth blocks have perfectly matching autocovariance structures. If $k \neq 0$ and $\ell \neq 0$, this is a contradiction as the blocks were defined such that two distinct blocks always correspond to differing autocovariance structures. Moreover, if either $k = 0$ or $\ell = 0$, then $\lambda_{\tau k} = \lambda_{\tau \ell} = 0$ and we have found a signal (block) that has all autocovariances zero, contradicting Assumption 1. Consequently, exactly one subvector of $\mathbf{w}$ is non-zero. As the choice of $\mathbf{w}$ within $\mathbf{W}$ was arbitrary, the result holds for all columns of $\mathbf{W}$.

We next show that exactly p_j columns of $\mathbf{W}$ have non-zero jth subvector, $j = 0, 1, \ldots, v$. Again the proof is by contradiction. Pick an arbitrary $j = 0, 1, \ldots, v$ and assume that more than p_j columns of $\mathbf{W}$ are such that their non-zero part lies in the jth subvector. Then these subvectors form a collection of more than p_j linearly independent vectors of length p_j (the linear independence follows as $\mathbf{W}$ is invertible and as each of its columns has non-zero elements in exactly one of the subvectors). This is a contradiction as no sets of linearly independent vectors with cardinality greater than n exist in $\mathbb{R}^n$. Thus at most p_j columns of $\mathbf{W}$ have non-zero jth subvector. Since the choice of j was arbitrary, the conclusion holds for all $j = 0, 1, \ldots, v$ and we conclude that the size of the jth block must be exactly p_j. Ordering the columns now suitably shows that $\mathbf{W}$ must have the form (S1.5), proving the first part of the argument.

To see the sufficiency of the block diagonal form (S1.5) we first notice that any matrix $\mathbf{W}$ that can be column-permuted so that $\mathbf{WP}$ is of the form (S1.5) satisfies $\mathbf{D}_\tau \mathbf{W}\mathbf{P} = \mathbf{W}\mathbf{P}\mathbf{D}_\tau$, $\tau \in \mathcal{T}$. Thus,

$$\begin{aligned} g(\mathbf{W}) &= \sum_{\tau \in \mathcal{T}} \left\| \mathrm{diag}\left(\mathbf{W}^\top \mathbf{D}_\tau \mathbf{W} \right) \right\|^2 \\ &= \sum_{\tau \in \mathcal{T}} \left\| \mathrm{diag}\left(\mathbf{P}\mathbf{P}^\top \mathbf{W}^\top \mathbf{D}_\tau \mathbf{W}\mathbf{P}\mathbf{P}^\top \right) \right\|^2 \\ &= \sum_{\tau \in \mathcal{T}} \left\| \mathrm{diag}\left(\mathbf{P}\mathbf{D}_\tau \mathbf{P}^\top \right) \right\|^2 \\ &= \sum_{\tau \in \mathcal{T}} \sum_{j=1}^{v} \lambda_{\tau j}^2 p_j, \end{aligned}$$

and we see that any $\mathbf{W}$ that is column-permutable to the form (S1.5) achieves the maximum. The sufficiency in conjunction with the necessity now equals the Fisher consistency of the population level problem.

We next move to the sample properties of the sequence of SOBI-solutions $\hat{\mathbf{U}}$ and show the consistency of its "off-diagonal blocks". That is, we prove that any sequence of maximizers $\hat{\mathbf{U}}$ of $\hat{g}$ can be permuted such that the off-diagonal blocks satisfy $\hat{\mathbf{U}}_{ij} \rightarrow_p \mathbf{0}$.

Let the set of all $p \times p$ orthogonal matrices be denoted by $\mathcal{U}^p$. To temporarily get rid of the unidentifiability of the ordering of the columns, we work in a specific subset of $\mathcal{U}^p$.

$$\mathcal{U}_0 = \{ \mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_p) \in \mathcal{U}^p \mid \mathbf{n}^\top \mathbf{u}_1^2 \geq \cdots \geq \mathbf{n}^\top \mathbf{u}_p^2 \},$$

where $\mathbf{u}^2 \in \mathbb{R}^p$ is the vector of element-wise squares of $\mathbf{u} \in \mathbb{R}^p$ and $\mathbf{n} = (p, p-1, \ldots, 1)^\top$. All orthogonal matrices $\mathbf{U} \in \mathcal{U}^p$ may have their columns permuted such that the permuted matrix belongs to $\mathcal{U}_0$. In case of ties in the condition defining $\mathcal{U}_0$, we arbitrarily choose one of the permutations. Let then $\hat{\mathbf{U}}$ be an arbitrary sequence of maximizers of $\hat{g}$, every term of which we assume, without loss of generality, to be a member of $\mathcal{U}_0$.

We first note that the uniform convergence of the sample objective function to the population objective function,

$$\sup_{\mathbf{U} \in \mathcal{U}_0} |\hat{g}(\mathbf{U}) - g(\mathbf{U})| \rightarrow_p 0, \tag{S1.7}$$

can be seen to hold as in the proof of (Miettinen et al., 2016, Theorem 1).

Let the set of all $\mathbf{U} \in \mathcal{U}^p$ of the form (S1.5) be denoted by $\mathcal{U}_P$ and define the set of all population level SOBI-solutions in $\mathcal{U}_0$ as

$$\mathcal{U}_S = \{\mathbf{U} \in \mathcal{U}_0 \mid g(\mathbf{U}) \geq g(\mathbf{V}), \text{ for all } \mathbf{V} \in \mathcal{U}_0\}.$$

We now claim that the set $\mathcal{U}_0$ is constructed such that we have $\mathcal{U}_S \subset \mathcal{U}_P$. To see this, we prove the contrapositive claim that $\mathcal{U} \setminus \mathcal{U}_P \subset \mathcal{U} \setminus \mathcal{U}_S$. Take an arbitrary $\mathbf{U} \in \mathcal{U} \setminus \mathcal{U}_P$. If $\mathbf{U}$ is not a maximizer of g, then clearly $\mathbf{U} \in \mathcal{U} \setminus \mathcal{U}_S$ and we are done. If instead $\mathbf{U}$ is a maximizer of g, then it must have two columns $\mathbf{u}_k, \mathbf{u}_\ell$ such that $k < \ell$ and $\mathbf{u}_k$ belongs to the ith column block and $\mathbf{u}_\ell$ belongs to the jth column block with $i > j$ (the two columns are in wrong order with respect to $\mathcal{U}_P$). However, then $\mathbf{n}^\top \mathbf{u}_k^2 \leq p - \sum_{k=1}^{i-1} p_k < p - \sum_{k=1}^{j} p_k + 1 \leq \mathbf{n}^\top \mathbf{u}_\ell^2$ and $\mathbf{U} \notin \mathcal{U}_0$, implying that again $\mathbf{U} \in \mathcal{U} \setminus \mathcal{U}_S$. Us having exhausted all cases, any $\mathbf{U} \in \mathcal{U}_S$ is thus also a member of $\mathcal{U}_P$ and has $\mathbf{U}_{ij} = 0$ for all $i \neq j$ where the partitioning is as in the statement of the lemma.

We prove the consistency via showing that the sequence of solutions $\hat{\mathbf{U}}$ gets arbitrarily close to the solution set $\mathcal{U}_S$ in the sense that,

$$\mathbb{P}(\inf_{\mathbf{V} \in \mathcal{U}_S} \|\hat{\mathbf{U}} - \mathbf{V}\|^2 > \varepsilon) \to 0, \quad \forall \varepsilon > 0.$$

To see this, fix $\varepsilon > 0$ and define the ε-neighbourhood of $\mathcal{U}_S$ in $\mathcal{U}_0$ as

$$\mathcal{U}_{S\varepsilon} = \{\mathbf{U} \in \mathcal{U}_0 \mid \inf_{\mathbf{V} \in \mathcal{U}_S} \|\mathbf{U} - \mathbf{V}\|^2 \leq \varepsilon\}.$$

Then

$$\mathbb{P}(\inf_{\mathbf{V} \in \mathcal{U}_S} \|\hat{\mathbf{U}} - \mathbf{V}\|^2 > \varepsilon) = \mathbb{P}(\hat{\mathbf{U}} \in \mathcal{U}_0 \setminus \mathcal{U}_{S\varepsilon}).$$

As all maximizers of g in $\mathcal{U}_0$ lie in $\mathcal{U}_S$, there exists $\delta = \delta(\varepsilon) > 0$ strictly positive such that for all $\mathbf{V} \in \mathcal{U}_0 \setminus \mathcal{U}_{S\varepsilon}$ we have $g(\mathbf{V}) < g(\mathbf{U}_S) - \delta$ where $\mathbf{U}_S$ is an arbitrary element of $\mathcal{U}_S$. This gives us,

$$\mathbb{P}(\inf_{\mathbf{V} \in \mathcal{U}_S} \|\hat{\mathbf{U}} - \mathbf{V}\|^2 > \varepsilon) \leq \mathbb{P}(g(\mathbf{U}_S) - g(\hat{\mathbf{U}}) > \delta).$$

By the definition of $\hat{\mathbf{U}}$ as a maximizer of $\hat{g}$, we have $\hat{g}(\hat{\mathbf{U}}) \geq \hat{g}(\mathbf{U}_S)$ and can construct the sequence of inequalities,

$$\begin{aligned} 0 &\leq g(\mathbf{U}_S) - g(\hat{\mathbf{U}}) \\ &\leq \hat{g}(\hat{\mathbf{U}}) - g(\hat{\mathbf{U}}) + g(\mathbf{U}_S) - \hat{g}(\mathbf{U}_S) \\ &\leq 2 \sup_{\mathbf{U}\in\mathcal{U}_0} |\hat{g}(\mathbf{U}) - g(\mathbf{U})|, \end{aligned}$$

where invoking (S1.7) shows that $g(\mathbf{U}_S) - g(\hat{\mathbf{U}}) \to_p 0$. Consequently,

$$\mathbb{P}(\inf_{\mathbf{V}\in\mathcal{U}_S} \|\hat{\mathbf{U}} - \mathbf{V}\|^2 > \varepsilon) \leq \mathbb{P}(g(\mathbf{U}_S) - g(\hat{\mathbf{U}}) > \delta) \to 0,$$

and we have that $\inf_{\mathbf{V}\in\mathcal{U}_S} \|\hat{\mathbf{U}} - \mathbf{V}\|^2 = o_p(1)$. Writing this block-wise and remembering that all elements of $\mathcal{U}_S \subset \mathcal{U}_P$ have off-diagonal blocks equal to zero, we get,

$$\inf_{\mathbf{V}\in\mathcal{U}_S} \|\hat{\mathbf{U}} - \mathbf{V}\|^2 = \inf_{\mathbf{V}\in\mathcal{U}_S} \left(\sum_{i=0}^{v} \|\hat{\mathbf{U}}_{ii} - \mathbf{V}_{ii}\|^2 + \sum_{i\neq j} \|\hat{\mathbf{U}}_{ij}\|^2 \right) \geq \sum_{i\neq j} \|\hat{\mathbf{U}}_{ij}\|^2,$$

implying that all off-diagonal blocks of $\hat{\mathbf{U}}$ satisfy $\|\hat{\mathbf{U}}_{ij}\| = o_p(1)$. Consequently, for every arbitrary sequence of solutions $\hat{\mathbf{U}}$, there exists a sequence of permutation matrices $\hat{\mathbf{P}}$ (chosen so that $\hat{\mathbf{U}}\hat{\mathbf{P}} \in \mathcal{U}_0$) such that the off-diagonal blocks of $\hat{\mathbf{U}}\hat{\mathbf{P}}$ converge in probability to zero.

2. Convergence rate

We next establish that the off-diagonal blocks of any sequence of solutions $\hat{\mathbf{U}} \in \mathcal{U}_0$ converge at the rate of root-T. The claimed result then follows for an arbitrary sequence of solutions $\hat{\mathbf{U}}$ by choosing the sequence of permutations $\hat{\mathbf{P}}$ such that $\hat{\mathbf{U}}\hat{\mathbf{P}} \in \mathcal{U}_0$.

By (Miettinen et al., 2016, Definition 2), the estimating equations of the SOBI-solution $\hat{\mathbf{U}} = (\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_p)$ are,

$$\mathbf{u}_k^\top \sum_{\tau \in \mathcal{T}} \hat{\mathbf{H}}_\tau \hat{\mathbf{u}}_\ell \mathbf{u}_\ell^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{u}}_\ell = \mathbf{u}_\ell^\top \sum_{\tau \in \mathcal{T}} \hat{\mathbf{H}}_\tau \hat{\mathbf{u}}_k \mathbf{u}_k^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{u}}_k, \quad \forall k, \ell = 1, \ldots, p, \tag{S1.8}$$

along with the orthogonality constraint $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_p$. The set of estimating equations (S1.8) can be written in matrix form as,

$$\sum_{\tau \in \mathcal{T}} \hat{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{U}} \mathrm{diag}(\hat{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{U}}) = \sum_{\tau \in \mathcal{T}} \mathrm{diag}(\hat{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{U}}) \hat{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{U}},$$

which is equivalent to claiming that the matrix $\hat{\mathbf{Y}} = \sum_{\tau \in \mathcal{T}} \hat{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{U}} \mathrm{diag}(\hat{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{U}})$ is symmetric, $\hat{\mathbf{Y}} = \hat{\mathbf{Y}}^\top$.

We next take $\hat{\mathbf{Y}}$, multiply it by $\sqrt{T}$ and expand as $\hat{\mathbf{H}}_\tau = \hat{\mathbf{H}}_\tau - \mathbf{D}_\tau + \mathbf{D}_\tau$ to obtain,

$$\begin{aligned}
\sqrt{T}\hat{\mathbf{Y}} =& \sum_{\tau \in \mathcal{T}} \hat{\mathbf{U}}^\top \sqrt{T}(\hat{\mathbf{H}}_\tau - \mathbf{D}_\tau) \hat{\mathbf{U}} \mathrm{diag}(\hat{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{U}}) \\
&+ \sum_{\tau \in \mathcal{T}} \hat{\mathbf{U}}^\top \mathbf{D}_\tau \hat{\mathbf{U}} \mathrm{diag}(\hat{\mathbf{U}}^\top \sqrt{T}(\hat{\mathbf{H}}_\tau - \mathbf{D}_\tau) \hat{\mathbf{U}}) \qquad \text{(S1.9)} \\
&+ \sqrt{T} \sum_{\tau \in \mathcal{T}} \hat{\mathbf{U}}^\top \mathbf{D}_\tau \hat{\mathbf{U}} \mathrm{diag}(\hat{\mathbf{U}}^\top \mathbf{D}_\tau \hat{\mathbf{U}}).
\end{aligned}$$

As $\hat{\mathbf{U}} = \mathcal{O}_p(1)$ by its orthogonality and $\sqrt{T}(\hat{\mathbf{H}}_\tau - \mathbf{D}_\tau) = \mathcal{O}_p(1)$ by Lemma 1, the first two terms on the right-hand side of (S1.9) are bounded in probability and we may lump them under a single $\mathcal{O}_p(1)$-symbol,

$$\sqrt{T}\hat{\mathbf{Y}} = \sqrt{T} \sum_{\tau \in \mathcal{T}} \hat{\mathbf{U}}^\top \mathbf{D}_\tau \hat{\mathbf{U}} \mathrm{diag}(\hat{\mathbf{U}}^\top \mathbf{D}_\tau \hat{\mathbf{U}}) + \mathcal{O}_p(1). \tag{S1.10}$$

Inspect next the term $\hat{\mathbf{D}}_\tau = \mathrm{diag}(\hat{\mathbf{U}}^\top \mathbf{D}_\tau \hat{\mathbf{U}})$. Performing the matrix multiplication block-wise we get as the (i,j)th block of $\hat{\mathbf{U}}^\top \mathbf{D}_\tau \hat{\mathbf{U}}$,

$$(\hat{\mathbf{U}}^\top \mathbf{D}_\tau \hat{\mathbf{U}})_{ij} = \sum_{k=0}^{v} \lambda_{\tau k} \hat{\mathbf{U}}_{ki}^\top \hat{\mathbf{U}}_{kj}.$$

As $\hat{\mathbf{U}}_{ij}^\top \to_p \mathbf{0}$ and $\hat{\mathbf{U}}_{ii}^\top \hat{\mathbf{U}}_{ii} \to_p \mathbf{I}_{p_i}$ (the latter follows from the orthogonality of $\hat{\mathbf{U}}$ and the consistency of its off-diagonal blocks), we have,

$$(\hat{\mathbf{U}}^\top \mathbf{D}_\tau \hat{\mathbf{U}})_{ij} = \delta_{ij} \lambda_{\tau i} \mathbf{I}_{p_i} + o_p(1),$$

where $\delta_{\cdot\cdot}$ is the Kronecker delta. Consequently,

$$\hat{\mathbf{D}}_\tau = \mathrm{diag}(\hat{\mathbf{U}}^\top \mathbf{D}_\tau \hat{\mathbf{U}}) = \mathbf{D}_\tau + o_p(1).$$

Denote by $\hat{\mathbf{U}}_{i,-j} \in \mathbb{R}^{(p-p_j)\times p_i}$ the ith column block of $\hat{\mathbf{U}}$ with the jth block removed, by $\mathbf{D}_{\tau,-j} \in \mathbb{R}^{(p-p_j)\times(p-p_j)}$ the result of removing the jth column and row blocks of $\mathbf{D}_\tau$ and by $\hat{\mathbf{D}}_{\tau j} \to_p \lambda_{\tau j} \mathbf{I}_{p_j}$ the jth $p_j \times p_j$ diagonal block of $\hat{\mathbf{D}}_\tau$. Our main claim is equivalent to requiring that,

$$\hat{\mathbf{U}}_{j,-j} = \mathcal{O}_p\left(\frac{1}{\sqrt{T}}\right), \quad \text{for all } j = 0, \ldots, v.$$

To show this, fix next j and take the (i,j)th block of the matrix $\sqrt{T}\hat{\mathbf{Y}}$ where $i \neq j$ is arbitrary and separate the jth term in the block-wise matrix multiplication of (S1.10) to obtain,

$$\sqrt{T}\hat{\mathbf{Y}}_{ij} = \sqrt{T} \sum_{\tau\in\mathcal{T}} \hat{\mathbf{U}}_{i,-j}^\top \mathbf{D}_{\tau,-j} \hat{\mathbf{U}}_{j,-j} \hat{\mathbf{D}}_{\tau j} + \sqrt{T} \sum_{\tau\in\mathcal{T}} \lambda_{\tau j} \hat{\mathbf{U}}_{ji}^\top \hat{\mathbf{U}}_{jj} \hat{\mathbf{D}}_{\tau j} + \mathcal{O}_p(1). \tag{S1.11}$$

Opening up the (i,j)th block (still with distinct i,j) of the orthogonality constraint $\hat{\mathbf{U}}^\top \hat{\mathbf{U}} = \mathbf{I}_p$ and again separating the jth term lets us write,

$$\hat{\mathbf{U}}_{ji}^\top \hat{\mathbf{U}}_{jj} = -\hat{\mathbf{U}}_{i,-j}^\top \hat{\mathbf{U}}_{j,-j}.$$

Plugging this in to (S1.11) gives us,

$$\sqrt{T}\hat{\mathbf{Y}}_{ij} = \sqrt{T}\sum_{\tau \in \mathcal{T}} \hat{\mathbf{U}}_{i,-j}^\top \mathbf{D}_{\tau,-j} \hat{\mathbf{U}}_{j,-j} \hat{\mathbf{D}}_{\tau j} - \sqrt{T}\sum_{\tau \in \mathcal{T}} \lambda_{\tau j} \hat{\mathbf{U}}_{i,-j}^\top \hat{\mathbf{U}}_{j,-j} \hat{\mathbf{D}}_{\tau j} + \mathcal{O}_p(1). \tag{S1.12}$$

Next we invoke the symmetry form, $\sqrt{T}\hat{\mathbf{Y}} = \sqrt{T}\hat{\mathbf{Y}}^\top$, of the estimating equations (S1.8). In block form the equations claim that $\sqrt{T}\hat{\mathbf{Y}}_{ij} = \sqrt{T}(\hat{\mathbf{Y}}_{ji})^\top$. Performing now the expansion equivalent to (S1.12) also for $\sqrt{T}(\hat{\mathbf{Y}}_{ji})^\top$ (again separating the jth block in the summation) and plugging in the expansions into the symmetry relation, we obtain,

$$\begin{aligned} \mathcal{O}_p(1) =& \sqrt{T}\sum_{\tau \in \mathcal{T}} \hat{\mathbf{U}}_{i,-j}^\top \mathbf{D}_{\tau,-j} \hat{\mathbf{U}}_{j,-j} \hat{\mathbf{D}}_{\tau j} - \sqrt{T}\sum_{\tau \in \mathcal{T}} \lambda_{\tau j} \hat{\mathbf{U}}_{i,-j}^\top \hat{\mathbf{U}}_{j,-j} \hat{\mathbf{D}}_{\tau j} \\ &- \sqrt{T}\sum_{\tau \in \mathcal{T}} \hat{\mathbf{D}}_{\tau i} \hat{\mathbf{U}}_{i,-j}^\top \mathbf{D}_{\tau,-j} \hat{\mathbf{U}}_{j,-j} + \sqrt{T}\sum_{\tau \in \mathcal{T}} \lambda_{\tau j} \hat{\mathbf{D}}_{\tau i} \hat{\mathbf{U}}_{i,-j}^\top \hat{\mathbf{U}}_{j,-j}. \end{aligned} \tag{S1.13}$$

We then pre-multiply (S1.13) by $\hat{\mathbf{U}}_{i,-j} = \mathcal{O}_p(1)$ and sum the result over the index $i \in \{0, \ldots, v\} \setminus \{j\}$. Denoting $\hat{\mathbf{A}}_i = \hat{\mathbf{U}}_{i,-j}$ this gives us,

$$\begin{aligned} \mathcal{O}_p(1) =& \sqrt{T}\sum_{\tau \in \mathcal{T}} \sum_{i \neq j} \hat{\mathbf{A}}_i \hat{\mathbf{A}}_i^\top \mathbf{D}_{\tau,-j} \hat{\mathbf{A}}_j \hat{\mathbf{D}}_{\tau j} - \sqrt{T}\sum_{\tau \in \mathcal{T}} \sum_{i \neq j} \lambda_{\tau j} \hat{\mathbf{A}}_i \hat{\mathbf{A}}_i^\top \hat{\mathbf{A}}_j \hat{\mathbf{D}}_{\tau j} \\ &- \sqrt{T}\sum_{\tau \in \mathcal{T}} \sum_{i \neq j} \hat{\mathbf{A}}_i \hat{\mathbf{D}}_{\tau i} \hat{\mathbf{A}}_i^\top \mathbf{D}_{\tau,-j} \hat{\mathbf{A}}_j + \sqrt{T}\sum_{\tau \in \mathcal{T}} \sum_{i \neq j} \lambda_{\tau j} \hat{\mathbf{A}}_i \hat{\mathbf{D}}_{\tau i} \hat{\mathbf{A}}_i^\top \hat{\mathbf{A}}_j. \end{aligned} \tag{S1.14}$$

We next row-vectorize (S1.14) to obtain us,

$$\mathcal{O}_p(1) = \sum_{\tau\in\mathcal{T}}\sum_{i\neq j}\Big[\hat{\mathbf{A}}_i\hat{\mathbf{A}}_i^\top\mathbf{D}_{\tau,-j}\otimes\hat{\mathbf{D}}_{\tau j} - \lambda_{\tau j}\hat{\mathbf{A}}_i\hat{\mathbf{A}}_i^\top\otimes\hat{\mathbf{D}}_{\tau j}$$
$$-\ \hat{\mathbf{A}}_i\hat{\mathbf{D}}_{\tau i}\hat{\mathbf{A}}_i^\top\mathbf{D}_{\tau,-j}\otimes\mathbf{I}_{p_j} + \lambda_{\tau j}\hat{\mathbf{A}}_i\hat{\mathbf{D}}_{\tau i}\hat{\mathbf{A}}_i^\top\otimes\mathbf{I}_{p_j}\Big]\sqrt{T}\text{vect}(\hat{\mathbf{A}}_j). \tag{S1.15}$$

By the consistency of the off-diagonal blocks of $\hat{\mathbf{U}}$, we have $\hat{\mathbf{U}}_{ij}\to_p\mathbf{0}$ for all $i\neq j$ and $\hat{\mathbf{U}}_{ii}\hat{\mathbf{U}}_{ii}^\top\to_p\mathbf{I}_{p_i}$ for all i. Consequently, we have the following convergences in probability, $\sum_{i\neq j}\hat{\mathbf{A}}_i\hat{\mathbf{A}}_i^\top\to_p\mathbf{I}_{p-p_j}$, $\sum_{i\neq j}\hat{\mathbf{A}}_i\hat{\mathbf{D}}_{\tau i}\hat{\mathbf{A}}_i^\top\to_p\mathbf{D}_{\tau,-j}$ and $\hat{\mathbf{D}}_{\tau j}\to_p\lambda_{\tau j}\mathbf{I}_{p_j}$. Calling next the matrix in the square brackets on the right-hand side of (S1.15) by $\hat{\mathbf{C}}\in\mathbb{R}^{(p-p_j)p_j\times(p-p_j)p_j}$, the convergences imply that,

$$\hat{\mathbf{C}}\to_p\mathbf{C} = \sum_{\tau\in\mathcal{T}}\left[\lambda_{\tau j}\mathbf{D}_{\tau,-j}\otimes\mathbf{I}_{p_j} - \lambda_{\tau j}^2\mathbf{I}_{(p-p_j)p_j} - \mathbf{D}_{\tau,-j}^2\otimes\mathbf{I}_{p_j} + \lambda_{\tau j}\mathbf{D}_{\tau,-j}\otimes\mathbf{I}_{p_j}\right]. \tag{S1.16}$$

The matrix $\mathbf{C}$ in(S1.16) is a diagonal matrix and its diagonal is divided into v segments of lengths p_ip_j, $i\in\{0,\ldots,v\}\setminus\{j\}$. Each segment matches with the vectorization of the corresponding block $\hat{\mathbf{U}}_{ij}$ in the vectorized matrix $\text{vect}(\hat{\mathbf{A}}_j)=\text{vect}(\hat{\mathbf{U}}_{j,-j})$. All elements of the ith segment of the diagonal of $\mathbf{C}$ are equal to,

$$\sum_{\tau\in\mathcal{T}}\left(\lambda_{\tau j}\lambda_{\tau i} - \lambda_{\tau j}^2 - \lambda_{\tau i}^2 + \lambda_{\tau j}\lambda_{\tau i}\right) = -\sum_{\tau\in\mathcal{T}}(\lambda_{\tau i}-\lambda_{\tau j})^2 < 0,$$

where the inequality follows from our definition of the blocks such that they differ in their autocovariances for at least one lag $\tau \in \mathcal{T}$. Thus the matrix $\mathbf{C}$ is invertible and we may pre-multiply (S1.15) by $\hat{\mathbf{C}}^{-1}$ which is asymptotically well-defined. By Slutsky's theorem (for random matrices) we obtain,

$$\sqrt{T}\text{vect}(\hat{\mathbf{A}}_j) = \hat{\mathbf{C}}^{-1}\mathcal{O}_p(1) = \mathcal{O}_p(1). \tag{S1.17}$$

As the choice of the column block j was arbitrary, the result (S1.17) holds for all $\hat{\mathbf{A}}_j = \hat{\mathbf{U}}_{j,-j}$, concluding the proof of Lemma 2. □

Proof of Corollary 1. The jth diagonal block of the orthogonality constraint $\hat{\mathbf{U}}^\top \hat{\mathbf{U}} = \mathbf{I}_p$ reads,

$$\sum_{k \neq j} \hat{\mathbf{U}}_{kj}^\top \hat{\mathbf{U}}_{kj} = \mathbf{I}_{p_j} - \hat{\mathbf{U}}_{jj}^\top \hat{\mathbf{U}}_{jj},$$

where the left-hand side is by Lemma 2 of order $\mathcal{O}_p(1/T)$, giving the first claim. The second one follows in a similar way by starting with $\hat{\mathbf{U}}\hat{\mathbf{U}}^\top = \mathbf{I}_p$ instead. □

Proof of Lemma 3. Recall the definition of $\hat{m}_q$ as,

$$\hat{m}_q = \frac{1}{|\mathcal{T}| r^2} \sum_{\tau \in \mathcal{T}} \| \hat{\mathbf{W}}_q^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_q \|^2,$$

where $\hat{\mathbf{W}}_q$ contains the columns of the SOBI-solution that correspond to the smallest q sums of squared pseudo-eigenvalues $\sum_{\tau \in \mathcal{T}} \text{diag}(\hat{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{U}})^2$.

By Lemma 2, $\hat{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{U}} = \hat{\mathbf{P}} \tilde{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \tilde{\mathbf{U}} \hat{\mathbf{P}}^\top$ where $\tilde{\mathbf{U}}$ is the block-diagonal matrix on the right-hand side of Lemma 2. We derive an asymptotic expression for the ith diagonal block $\hat{\mathbf{E}}_{\tau ii}$ of the matrices $\hat{\mathbf{E}}_\tau = \tilde{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \tilde{\mathbf{U}}$. By Lemmas 1, 2, Corollary 1 and Assumption 3,

$$\begin{aligned} \hat{\mathbf{E}}_{\tau ii} &= \sum_{s=0}^{v} \sum_{t=0}^{v} \hat{\mathbf{U}}_{si}^\top \hat{\mathbf{H}}_{\tau st} \hat{\mathbf{U}}_{ti} \\ &= \sum_{s \neq t} \hat{\mathbf{U}}_{si}^\top \hat{\mathbf{H}}_{\tau st} \hat{\mathbf{U}}_{ti} + \sum_{s=0}^{v} \hat{\mathbf{U}}_{si}^\top (\hat{\mathbf{H}}_{\tau ss} - \lambda_{\tau s} \mathbf{I}_{p_s}) \hat{\mathbf{U}}_{si} + \sum_{s=0}^{v} \lambda_{\tau s} \hat{\mathbf{U}}_{si}^\top \hat{\mathbf{U}}_{si} \\ &= \lambda_{\tau i} \mathbf{I}_{p_s} + \hat{\mathbf{U}}_{ii}^\top (\hat{\mathbf{H}}_{\tau ii} - \lambda_{\tau i} \mathbf{I}_{p_i}) \hat{\mathbf{U}}_{ii} + \mathcal{O}_p(1/T), \end{aligned} \tag{S1.18}$$

where $\hat{\mathbf{H}}_{\tau st}$ is the (s,t)th block of $\hat{\mathbf{H}}_\tau$ in the indexing of Lemma 2 and $\lambda_{\tau j}$ denotes the autocovariance of the jth group for lag τ in the corresponding grouping. As $(\hat{\mathbf{H}}_{\tau ii} - \lambda_{\tau i} \mathbf{I}_{p_i}) = \mathcal{O}_p(1/\sqrt{T})$, we have by (S1.18) that the pseudo-eigenvalues converge in probability to the respective population values,

$$\sum_{\tau \in \mathcal{T}} \mathrm{diag}(\hat{\mathbf{E}}_\tau)^2 \to_p \sum_{\tau \in \mathcal{T}} \mathbf{D}_\tau^2. \tag{S1.19}$$

Let A_q denote the event that the last q columns of $\hat{\mathbf{U}}$ are up to ordering equal to the last q columns of $\tilde{\mathbf{U}}$, that is, the ordering based on the estimated sums of squared pseudo-eigenvalues correctly identifies the noise components. By Assumption 1, the signals are well-separated from the noise in the sense that no signal corresponds to the value zero in the diagonal of $\sum_{\tau \in \mathcal{T}} \mathbf{\Lambda}_\tau^2$

and consequently, by (S1.19), we have $\mathbb{P}(A_q) \to 1$.

Denote next the final column block of $\tilde{\mathbf{U}}$ by $\tilde{\mathbf{U}}_q \in \mathbb{R}^r$. Conditional on A_q, the two column blocks are the same up to a permutation, $\hat{\mathbf{W}}_q = \tilde{\mathbf{U}}_q \hat{\mathbf{P}}_q$ for some sequence of permutation matrices $\hat{\mathbf{P}}_q \in \mathbb{R}^{r \times r}$, and we can write for an arbitrary $\varepsilon > 0$,

$$
\begin{aligned}
&\mathbb{P}\left(\sqrt{T}\left|\|\hat{\mathbf{W}}_q^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_q\| - \|\tilde{\mathbf{U}}_q^\top \hat{\mathbf{H}}_\tau \tilde{\mathbf{U}}_q\|\right| < \varepsilon\right) \\
=&\mathbb{P}\left(\sqrt{T}\left|\|\hat{\mathbf{W}}_q^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_q\| - \|\tilde{\mathbf{U}}_q^\top \hat{\mathbf{H}}_\tau \tilde{\mathbf{U}}_q\|\right| < \varepsilon \mid A_q\right) \mathbb{P}(A_q) \\
+&\mathbb{P}\left(\sqrt{T}\left|\|\hat{\mathbf{W}}_q^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_q\| - \|\tilde{\mathbf{U}}_q^\top \hat{\mathbf{H}}_\tau \tilde{\mathbf{U}}_q\|\right| < \varepsilon \mid A_q^c\right) \mathbb{P}(A_q^c) \\
=&\mathbb{P}(A_q) + \mathbb{P}\left(\sqrt{T}\left|\|\hat{\mathbf{W}}_q^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_q\| - \|\tilde{\mathbf{U}}_q^\top \hat{\mathbf{H}}_\tau \tilde{\mathbf{U}}_q\|\right| < \varepsilon \mid A_q^c\right) (1 - \mathbb{P}(A_q)) \to 1,
\end{aligned}
$$

showing the convergence in probability, $\sqrt{T}\|\hat{\mathbf{W}}_q^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_q\| = \sqrt{T}\|\tilde{\mathbf{U}}_q^\top \hat{\mathbf{H}}_\tau \tilde{\mathbf{U}}_q\| + o_p(1)$, for all $\tau \in \mathcal{T}$. Furthermore, by (S1.18),

$$
\sqrt{T}\|\tilde{\mathbf{U}}_q^\top \hat{\mathbf{H}}_\tau \tilde{\mathbf{U}}_q\| = \sqrt{T}\|\hat{\mathbf{E}}_{\tau 00}\| = \|\sqrt{T}\hat{\mathbf{U}}_{00}^\top \hat{\mathbf{H}}_{\tau 00} \hat{\mathbf{U}}_{00} + \mathcal{O}_p(1/\sqrt{T})\| = \mathcal{O}_p(1),
$$

showing that,

$$
\begin{aligned}
T\|\hat{\mathbf{W}}_q^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_q\|^2 &= \|\sqrt{T}\hat{\mathbf{U}}_{00}^\top \hat{\mathbf{H}}_{\tau 00} \hat{\mathbf{U}}_{00} + \mathcal{O}_p(1/\sqrt{T})\|^2 + o_p(1) \\
&= \|\sqrt{T}\hat{\mathbf{U}}_{00}^\top \hat{\mathbf{H}}_{\tau 00} \hat{\mathbf{U}}_{00}\|^2 + o_p(1) \\
&= T \cdot \mathrm{tr}(\hat{\mathbf{U}}_{00} \hat{\mathbf{U}}_{00}^\top \hat{\mathbf{H}}_{\tau 00} \hat{\mathbf{U}}_{00} \hat{\mathbf{U}}_{00}^\top \hat{\mathbf{H}}_{\tau 00}) + o_p(1) \\
&= T\|\hat{\mathbf{H}}_{\tau 00}\|^2 + o_p(1) \\
&= T\|\hat{\mathbf{R}}_{\tau 00}\|^2 + o_p(1),
\end{aligned}
$$

where the second-to-last equality uses Corollary 1 and the last one Lemma 1. Substituting now into the definition of $\hat{m}_q$, we obtain the claim,

$$T \cdot \hat{m}_q = \frac{T}{|\mathcal{T}|r^2} \sum_{\tau \in \mathcal{T}} \|\hat{\mathbf{W}}_q^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_q\|^2 = \frac{T}{|\mathcal{T}|r^2} \sum_{\tau \in \mathcal{T}} \|\hat{\mathbf{R}}_{\tau 00}\|^2 + o_p(1).$$

□

Proof of Lemma 4. Write first,

$$\begin{aligned}\hat{\mathbf{S}}_\tau &= \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_{t+\tau} - \bar{\mathbf{x}})^\top \\ &= \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{x}_t \mathbf{x}_{t+\tau}^\top - \bar{\mathbf{x}} \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{x}_{t+\tau}^\top - \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{x}_t \bar{\mathbf{x}}^\top + \bar{\mathbf{x}}\bar{\mathbf{x}}^\top.\end{aligned}$$

By Assumption 2 and (Brockwell and Davis, 1991, Proposition 11.2.2), the latent series $\mathbf{z}_t = \mathbf{x}_t$ (we use identity mixing) satisfy a central limit theorem, implying that $\bar{\mathbf{x}} = \mathcal{O}_p(1/\sqrt{T})$. Thus,

$$\sqrt{T}(\hat{\mathbf{S}}_\tau - \mathbf{D}_\tau) = \sqrt{T}(\frac{1}{T-\tau} \sum_{t=1}^{T-\tau} \mathbf{x}_t \mathbf{x}_{t+\tau}^\top - \mathbf{D}_\tau) + \mathcal{O}_p(1/\sqrt{T}),$$

and it is sufficient to show the limiting result for the non-centered covariance and autocovariance matrices. Consequently, in the following we implicitly assume that no centering is used.

The blocks $\hat{\mathbf{R}}_{\tau_1 00}, \ldots, \hat{\mathbf{R}}_{\tau_{|\mathcal{T}|} 00}$ are the symmetrized autocovariance matrices of the white noise part of $\mathbf{z}_t$. By Assumption 2, the latent series $\mathbf{z}_t$ has an MA(∞)-representation and by considering only the last r components of the representation we see that also the white noise part has separately an MA(∞)-representation. Now the lower right diagonal blocks of the matrices $\boldsymbol{\Psi}_j$ take the roles of $\boldsymbol{\Psi}_j$ and by Assumption 2 these blocks equal $\boldsymbol{\Psi}_{j00} = \delta_{j0}\mathbf{I}_r$. Consequently, by (Miettinen et al., 2016, Lemma 1) the vector,

$$\sqrt{T}\text{vec}\left(\hat{\mathbf{R}}_{\tau_1 00}, \ldots, \hat{\mathbf{R}}_{\tau_{|\mathcal{T}|}00}\right),$$

admits a limiting multivariate normal distribution with zero mean and the covariance matrix equal to

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \cdots & \mathbf{V}_{1|\mathcal{T}|} \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{|\mathcal{T}|1} & \cdots & \mathbf{V}_{|\mathcal{T}||\mathcal{T}|} \end{pmatrix} \in \mathbb{R}^{|\mathcal{T}|r^2 \times |\mathcal{T}|r^2}, \tag{S1.20}$$

where $\mathbf{V}_{\ell m} = \text{diag}(\text{vec}(\mathbf{D}_{\ell m}))(\mathbf{K}_{rr} - \mathbf{D}_{rr} + \mathbf{I}_{r^2})$. The matrices $\mathbf{D}_{\ell m} \in \mathbb{R}^{r \times r}$, $\ell, m = 0, \ldots, |\mathcal{T}|$ (we do not use the zero index here but it appears in the following formulas so it is included), are defined element-wise as,

$$\begin{aligned} (\mathbf{D}_{\ell m})_{ii} &= (\beta_i - 3)(\mathbf{F}_\ell)_{ii}(\mathbf{F}_m)_{ii} + \sum_{k=-\infty}^{\infty} [(\mathbf{F}_{k+\ell})_{ii}(\mathbf{F}_{k+m})_{ii} + (\mathbf{F}_{k+\ell})_{ii}(\mathbf{F}_{k-m})_{ii}] \\ (\mathbf{D}_{\ell m})_{ij} &= (\beta_{ij} - 1)(\mathbf{F}_\ell + \mathbf{F}_\ell^\top)_{ij}(\mathbf{F}_m + \mathbf{F}_m^\top)_{ij} \\ &\quad + \frac{1}{2}\sum_{k=-\infty}^{\infty} [(\mathbf{F}_{k+\ell-m})_{ii}(\mathbf{F}_k)_{jj} + (\mathbf{F}_k)_{ii}(\mathbf{F}_{k+\ell+m})_{jj}], \quad i \neq j. \end{aligned} \tag{S1.21}$$

where $\beta_i = \text{E}(\epsilon_{ti}^4)$, $\beta_{ij} = \text{E}(\epsilon_{ti}^2\epsilon_{ti}^2)$ and ϵ_{ti}, $i = 1, \ldots, r$, refers to the ith innovation component in the MA(∞)-representation of the white noise part.

The matrices $\mathbf{F}_\ell$ are defined as $\mathbf{F}_\ell = \sum_{j=-\infty}^{\infty} \boldsymbol{\psi}_j \boldsymbol{\psi}_{j+\ell}^\top$ where the vectors $\boldsymbol{\psi}_j \in \mathbb{R}^r$ contain the diagonal elements of the matrices $\boldsymbol{\Psi}_{j00}$.

Under Assumption 2 we have $\boldsymbol{\psi}_j = \delta_{j0}\mathbf{1}_r$ where the vector $\mathbf{1}_r \in \mathbb{R}^r$ consists solely of zeroes. Consequently $\mathbf{F}_\ell = \delta_{\ell 0}\mathbf{J}_r$. Plugging this in to (S1.21) gives for the diagonal elements of $\mathbf{D}_{\ell m}$ that

$$\begin{aligned}
(\mathbf{D}_{\ell m})_{ii} &= (\beta_i - 3)\delta_{\ell 0}\delta_{m0} + \sum_{k=-\infty}^{\infty} \left[\delta_{(k+\ell)0}\delta_{(k+m)0} + \delta_{(k+\ell)0}\delta_{(k-m)0}\right] \\
&= (\beta_i - 3)\delta_{\ell 0}\delta_{m0} + \delta_{\ell m} + \delta_{\ell 0}\delta_{m0},
\end{aligned}$$

which implies that the matrices $\mathbf{V}_{\ell m}$, $\ell, m = 1, \ldots, |\mathcal{T}|$, have non-zero diagonals precisely when $\ell = m$ and then the diagonal is filled with ones. Plugging $\mathbf{F}_\ell = \delta_{\ell 0}\mathbf{J}_r$ in to the definition of the diagonal elements in (S1.21) gives,

$$\begin{aligned}
(\mathbf{D}_{\ell m})_{ij} &= (\beta_{ij} - 1)2\delta_{\ell 0}2\delta_{\ell 0} + \frac{1}{2}\sum_{k=-\infty}^{\infty} \left[\delta_{(k+\ell-m)0}\delta_{k0} + \delta_{k0}\delta_{(k+\ell+m)0}\right] \\
&= 4(\beta_{ij} - 1)\delta_{\ell 0}\delta_{\ell 0} + \frac{1}{2}(\delta_{\ell m} + \delta_{\ell 0}\delta_{m0}),
\end{aligned}$$

which says that the matrices $\mathbf{V}_{\ell m}$, $\ell, m = 1, \ldots, |\mathcal{T}|$, have non-zero off-diagonals precisely when $\ell = m$ and then the off-diagonal is filled with one-halves.

Combining the forms for the diagonals and off-diagonals, we get $\mathbf{D}_{\ell m} = \delta_{\ell m}(\mathbf{J}_r + \mathbf{I}_r)/2$, $\ell, m = 1, \ldots, |\mathcal{T}|$. Plugging this in to (S1.20) now gives the claim and concludes the proof.

□

Proof of Lemma 5. The case $k = q$ follows immediately from Proposition 1.

For the case $k > q$, we note that,

$$|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k = \sum_{\tau \in \mathcal{T}} \|\hat{\mathbf{W}}_k^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_k\|^2 \leq \sum_{\tau \in \mathcal{T}} \|\hat{\mathbf{W}}_q^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_q\|^2 = |\mathcal{T}|(p-q)^2 \cdot \hat{m}_q,$$

as $\hat{\mathbf{W}}_k^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_k$ is a sub-matrix of $\hat{\mathbf{W}}_q^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_q$. The case $k > q$ now follows from the case $k = q$ and the positivity of the test statistic.

Finally, for the case $k < q$, we write,

$$|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k = \sum_{\tau \in \mathcal{T}} \|\hat{\mathbf{W}}_k^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{W}}_k\|^2 \geq \sum_{\tau \in \mathcal{T}} (\hat{\mathbf{v}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{v}})^2, \tag{S1.22}$$

where $\hat{\mathbf{v}}$ is the column of $\hat{\mathbf{U}}$ which corresponds to the qth largest diagonal element of $\sum_{\tau \in \mathcal{T}} \mathrm{diag}(\hat{\mathbf{U}}^\top \hat{\mathbf{H}}_\tau \hat{\mathbf{U}})^2$ (the final presumed signal component).

Denote the p_v columns of the block matrix on the right-hand side of Lemma 2 that are associated with the vth signal group by $\tilde{\mathbf{v}}_1, \ldots, \tilde{\mathbf{v}}_{p_v}$. For large values of T we expect the vector $\hat{\mathbf{v}}$ to equal one of these vectors with high probability. Indeed, denoting $g(\mathbf{v}) = \sum_{\tau \in \mathcal{T}} (\mathbf{v}^\top \hat{\mathbf{H}}_\tau \mathbf{v})^2$, we can show using the conditional probability trick of Lemma 3 that

$$\min_{j=1,\ldots,p_v} |g(\hat{\mathbf{v}}) - g(\tilde{\mathbf{v}}_j)| = o_p(1).$$

By Lemmas 1, 2 and Corollary 1 we additionally have for all $j = 1, \ldots, p_v$

that $g(\tilde{\mathbf{v}}_j) \to_p b = \sum_{\tau\in\mathcal{T}} \lambda_{\tau v}^2$, where $\lambda_{\tau v}$ are the autocovariances of the final signal group. By Assumption 1, moreover $b > 0$.

We then bound,

$$|g(\hat{\mathbf{v}}) - b| \le |g(\hat{\mathbf{v}}) - g(\tilde{\mathbf{v}}_j)| + |g(\tilde{\mathbf{v}}_j) - b| \le |g(\hat{\mathbf{v}}) - g(\tilde{\mathbf{v}}_j)| + \sum_{j'=1}^{p_v} |g(\tilde{\mathbf{v}}_{j'}) - b|\,,$$

which holds for all $j = 1, \ldots, p_v$, and consequently,

$$|g(\hat{\mathbf{v}}) - b| \le \min_{j=1,\ldots,p_v} |g(\hat{\mathbf{v}}) - g(\tilde{\mathbf{v}}_j)| + \sum_{j'=1}^{p_v} |g(\tilde{\mathbf{v}}_{j'}) - b| = o_p(1).$$

This implies that $g(\hat{\mathbf{v}}) \to_p b > 0$, which, when plugged in into (S1.22), yields the final part of the claim. □

Proof of Proposition 1. By Lemma 3, the limiting distribution of $T|\mathcal{T}|r^2 \cdot \hat{m}_q$ is the same as the limiting distribution of

$$\begin{aligned} T|\mathcal{T}|r^2 \cdot \hat{m}_q^* &= \sum_{\tau\in\mathcal{T}} \|\sqrt{T}\hat{\mathbf{R}}_{\tau 00}\|^2 \\ &= \|\sqrt{T}\mathrm{vec}\left(\hat{\mathbf{R}}_{\tau_1 00}, \ldots, \hat{\mathbf{R}}_{\tau_{|\mathcal{T}|} 00}\right)\|^2 \\ &= \sqrt{T}\mathrm{vec}^\top\left(\hat{\mathbf{R}}_{\tau_1 00}, \ldots, \hat{\mathbf{R}}_{\tau_{|\mathcal{T}|} 00}\right)\sqrt{T}\mathrm{vec}\left(\hat{\mathbf{R}}_{\tau_1 00}, \ldots, \hat{\mathbf{R}}_{\tau_{|\mathcal{T}|} 00}\right). \end{aligned}$$

By Lemma 4 and the continuous mapping theorem, the limiting distribution of $T|\mathcal{T}|r^2 \cdot \hat{m}_q^*$ is the same as the distribution of $\mathbf{y}^\top\mathbf{y}$ where $\mathbf{y}$ is a mean-zero multivariate normal random vector with the covariance matrix $\mathbf{V}$ given in Lemma 4. Equivalently, the limiting distribution of $T|\mathcal{T}|r^2 \cdot \hat{m}_q^*$ is the same as the distribution of $\mathbf{y}_0^\top\mathbf{V}\mathbf{y}_0$ where $\mathbf{y}_0$ is a standardized multivariate normal random vector. By (Serfling, 2009, Chapter 3.5), if $\mathbf{V}$ is idempotent and symmetric, then the limiting distribution of $\mathbf{y}_0^\top \mathbf{V} \mathbf{y}_0$ is $\chi^2_{\mathrm{tr}(\mathbf{V})}$. To see that $\mathbf{V}$ is indeed idempotent, we inspect the square of its arbitrary diagonal block $\mathbf{V}_0$,

$$\mathbf{V}_0^2 = \left[\mathrm{diag}(\mathrm{vec}(\mathbf{J}_r + \mathbf{I}_r)/2)(\mathbf{K}_{rr} - \mathbf{D}_{rr} + \mathbf{I}_{r^2})\right]^2.$$

We simplify using $\mathrm{diag}(\mathrm{vec}(\mathbf{J}_r)) = \mathbf{I}_{r^2}$, $\mathrm{diag}(\mathrm{vec}(\mathbf{I}_r)) = \mathbf{D}_{rr}$, $\mathbf{D}_{rr}\mathbf{K}_{rr} = \mathbf{D}_{rr}$, $\mathbf{D}_{rr}^2 = \mathbf{D}_{rr}$ and $\mathbf{K}_{rr}^2 = \mathbf{I}_{r^2}$, to obtain $\mathbf{V}_0 = (\mathbf{K}_{rr} + \mathbf{I}_{r^2})/2$, which is symmetric, and,

$$\mathbf{V}_0^2 = \left[\frac{1}{2}(\mathbf{K}_{rr} + \mathbf{I}_{r^2})\right]^2 = \frac{1}{4}(2\mathbf{K}_{rr} + 2\mathbf{I}_{r^2}) = \mathbf{V}_0.$$

Thus $\mathbf{V}_0$ is idempotent and symmetric and consequently $\mathbf{V}$, constituting solely of the $|\mathcal{T}|$ diagonal blocks each equal to $\mathbf{V}_0$, is also idempotent and symmetric. The trace of $\mathbf{V}$ is $|\mathcal{T}|$ times the trace of $\mathbf{V}_0$, which equals,

$$\mathrm{tr}(\mathbf{V}_0) = \frac{1}{2}\mathrm{tr}(\mathbf{K}_{rr}) + \frac{1}{2}\mathrm{tr}(\mathbf{I}_{r^2}) = \frac{1}{2}(r + r^2) = \frac{1}{2}r(r+1).$$

The trace of $\mathbf{V}$ is then $|\mathcal{T}|r(r+1)/2$ and we have proved that the limiting distribution of $\mathbf{y}_0^\top \mathbf{V} \mathbf{y}_0$, and consequently that of $T|\mathcal{T}|r^2 \cdot \hat{m}_q$, is $\chi^2_{|\mathcal{T}|r(r+1)/2}$.

□

Proof of Proposition 2. Fix an arbitrary $\varepsilon > 0$. We want to show that $\mathbb{P}(|\hat{q} - q| < \varepsilon) \to 1$. Denote by B_k the event that $T|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k < c_{k,T}$.

Then,

$$\begin{aligned}\mathbb{P}(|\hat{q}-q|<\varepsilon) \geq \mathbb{P}(\hat{q}=q) &= \mathbb{P}(B_q \cap B_{q-1}^c \cap B_{q-2}^c \cap \cdots \cap B_0^c) \\ &= \mathbb{P}\left[(B_q^c \cup B_{q-1} \cup B_{q-2} \cup \cdots \cup B_0)^c\right] \\ &= 1-\mathbb{P}(B_q^c \cup B_{q-1} \cup B_{q-2} \cup \cdots \cup B_0) \\ &\geq 1-\mathbb{P}(B_q^c)-\sum_{k=0}^{q-1}\mathbb{P}(B_k),\end{aligned}$$

where the final inequality uses the union bound. The desired result follows by showing that $\mathbb{P}(B_q^c) \to 0$ and $\mathbb{P}(B_k) \to 0$ for all $k = 0, \ldots, q-1$.

The first of these probabilities has the form,

$$\mathbb{P}(T|\mathcal{T}|(p-q)^2 \cdot \hat{m}_q \geq c_{q,T}), \tag{S1.23}$$

where the left-hand side of the inequality (denoted in the following by $\hat{h}_q$) is by Lemma 5 asymptotically $\mathcal{O}_p(1)$ and by our assumptions $c_{q,T} \to \infty$. As such, for an arbitrary $\delta > 0$, there exists $M > 0$ and T_0 such that $\mathbb{P}(\hat{h}_q > M) < \delta$ and $c_{q,T} > M$ for all $T > T_0$. Consequently, for all $T > T_0$, we have for the probability (S1.23) that

$$\mathbb{P}(T|\mathcal{T}|(p-q)^2 \cdot \hat{m}_q \geq c_{q,T}) \leq \mathbb{P}(\hat{h}_q > M) < \delta,$$

yielding $\mathbb{P}(B_q^c) \to 0$.

By Lemma 5, the probabilities $\mathbb{P}(B_k)$, $k = 0, \ldots, q-1$, satisfy,

$$\mathbb{P}\left(|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k < \frac{c_{k,T}}{T}\right) \leq \mathbb{P}\left(\hat{s} < \frac{c_{k,T}}{T}\right), \tag{S1.24}$$

where we denote $\hat{s} = b + o_p(1)$ and $c_{k,T}/T \to 0$ by our assumptions. Consequently, for an arbitrary $\delta > 0$, there exists T_0 such that for all $T > T_0$ we have $\mathbb{P}(|\hat{s} - b| > b/2) < \delta$ and $c_{k,T}/T < b/2$. Taking then $T > T_0$, these combine to give for the probability (S1.24) that,

$$\mathbb{P}\left(|\mathcal{T}|(p-k)^2 \cdot \hat{m}_k < \frac{c_{k,T}}{T}\right) \leq \mathbb{P}\left(\hat{s} < \frac{b}{2}\right) < \delta,$$

yielding $\mathbb{P}(B_k) \to 0$ for any $k < q$ and concluding the proof. □

S2. Mixed signals of the data example

The 20 mixed components of the data example in Section 5 are shown in Figure S1.

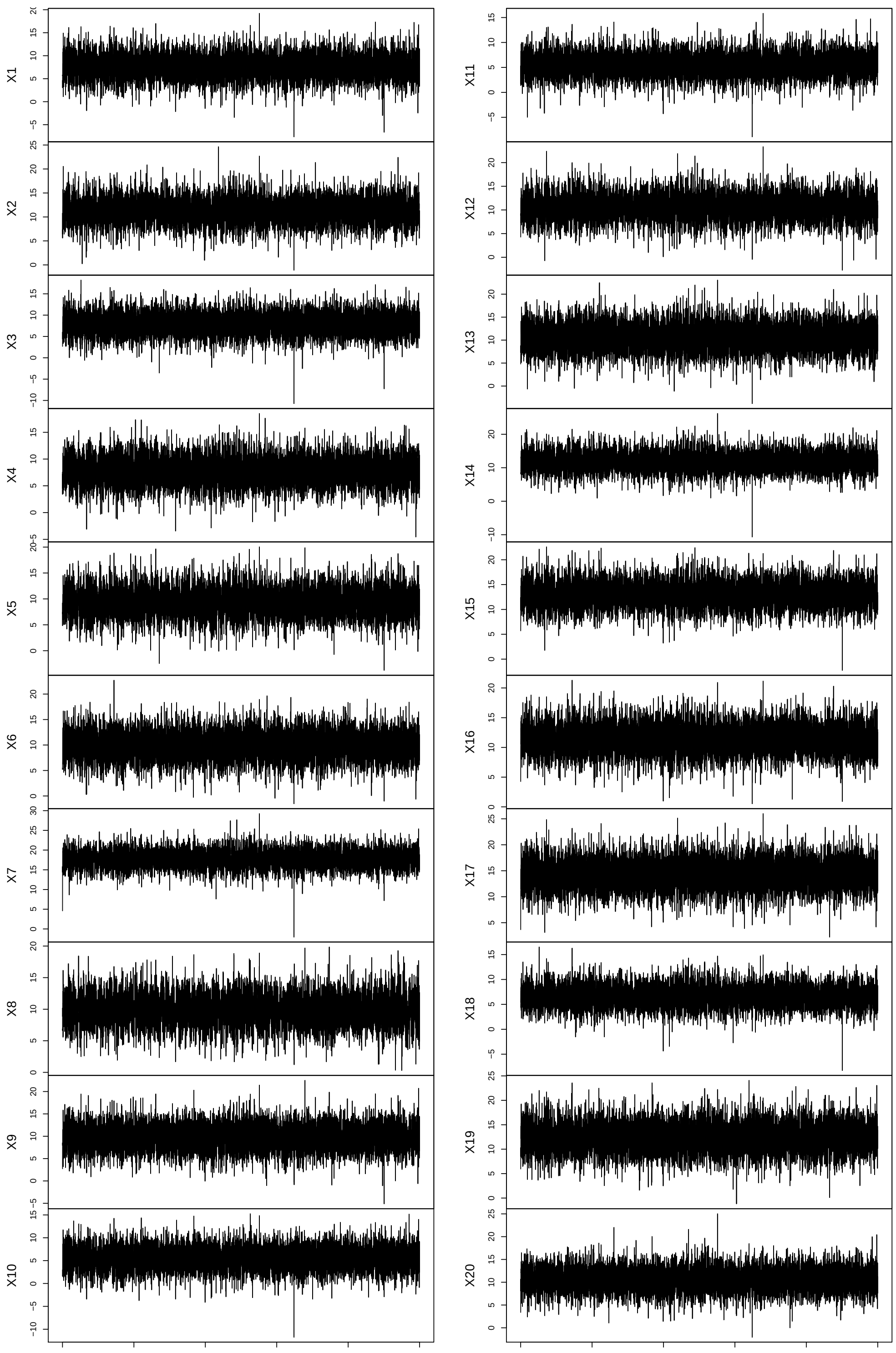

Figure S1: The 20-variate sound data time series.

S3. Supplementary simulations

The following subsections contain additional simulations to complement Sections 4.1 and 4.2. The result are in general quite similar to the ones shown in the main text, and only between Settings H3 and H3t are there any notable differences.

S3.1 Evaluation of the hypothesis testing under heavy tails

To evaluate the effect of non-Gaussian innovations and non-Gaussian white noise we modify settings H1–H3 from Section 4.1 as follows:

Setting H1t: MA(3), AR(2) and ARMA(1,1) having univariate t_5-distributed innovations together with two t_5-distributed white noise components.

Setting H2t: MA(10), MA(15) and M(20) processes having univariate t_5-distributed innovations together with two t_5-distributed white noise components.

Setting H3t: Three MA(3) processes having univariate t_5-distributed innovations and identical autocovariance functions together with two t_5-distributed white noise processes.

In all three cases the t_5-distributions were standardized to have unit variances.

Table S1: Rejection frequencies of H_{02} in Setting H1t at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	1.000	1.000	1.000	1.000	0.998	1.000
500	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000
5000	1.000	1.000	1.000	1.000	1.000	1.000

Table S2: Rejection frequencies of H_{03} in Setting H1t at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.052	0.050	0.079	0.054	0.103	0.056
500	0.051	0.052	0.060	0.048	0.066	0.050
1000	0.051	0.050	0.052	0.044	0.052	0.041
2000	0.043	0.046	0.046	0.050	0.050	0.054
5000	0.049	0.051	0.046	0.052	0.050	0.046

Table S3: Rejection frequencies of H_{04} in Setting H1t at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.002	0.008	0.016	0.005	0.026	0.008
500	0.008	0.010	0.009	0.006	0.010	0.002
1000	0.009	0.008	0.012	0.004	0.008	0.004
2000	0.004	0.008	0.006	0.004	0.011	0.004
5000	0.008	0.012	0.008	0.003	0.005	0.002

Table S4: Rejection frequencies of H_{02} in Setting H2t at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.039	0.040	0.594	0.473	0.908	0.845
500	0.098	0.098	0.980	0.982	1.000	1.000
1000	0.172	0.162	1.000	1.000	1.000	1.000
2000	0.272	0.270	1.000	1.000	1.000	1.000
5000	0.570	0.576	1.000	1.000	1.000	1.000

Table S5: Rejection frequencies of H_{03} in Setting H2t at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.002	0.006	0.124	0.047	0.142	0.053
500	0.004	0.009	0.078	0.046	0.084	0.056
1000	0.012	0.018	0.055	0.042	0.054	0.048
2000	0.017	0.022	0.053	0.052	0.060	0.058
5000	0.031	0.038	0.049	0.045	0.057	0.055

Table S6: Rejection frequencies of H_{04} in Setting H2t at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.000	0.002	0.036	0.005	0.038	0.006
500	0.000	0.006	0.012	0.004	0.014	0.004
1000	0.002	0.006	0.010	0.004	0.006	0.002
2000	0.003	0.004	0.008	0.005	0.014	0.006
5000	0.002	0.010	0.007	0.004	0.010	0.004

Table S7: Rejection frequencies of H_{02} in Setting H3t at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	1.000	1.000	1.000	1.000	0.996	0.993
500	1.000	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000	1.000
2000	1.000	1.000	1.000	1.000	1.000	1.000
5000	1.000	1.000	1.000	1.000	1.000	1.000

Table S8: Rejection frequencies of H_{03} in Setting H3t at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.058	0.054	0.082	0.043	0.134	0.058
500	0.051	0.047	0.059	0.047	0.073	0.054
1000	0.054	0.054	0.052	0.050	0.060	0.054
2000	0.040	0.044	0.060	0.047	0.058	0.060
5000	0.049	0.050	0.050	0.048	0.046	0.044

Table S9: Rejection frequencies of H_{04} in Setting H3t at level $\alpha = 0.05$.

	AMUSE		SOBI6		SOBI12	
n	Asymp	Boot	Asymp	Boot	Asymp	Boot
200	0.006	0.008	0.018	0.006	0.032	0.006
500	0.006	0.010	0.012	0.004	0.012	0.004
1000	0.005	0.008	0.008	0.004	0.011	0.006
2000	0.006	0.008	0.008	0.006	0.011	0.006
5000	0.005	0.006	0.006	0.002	0.007	0.002

S3.2 Evaluation of determining the dimension of the signal under heavy tails

To evaluate the effect of non-Gaussian innovations and non-Gaussian white noise we modify settings D1–D3 from Section 4.2 as follows:

Setting D1t: AR(2), AR(3), ARMA(1,1), ARMA(3,2) and MA(3) processes having univariate t_5-distributed innovations together with five t_5-distributed white noise components.

Setting D2t: Same processes as in D1t but the MA(3) is changed to an MA(1) process with the parameter equal to 0.1.

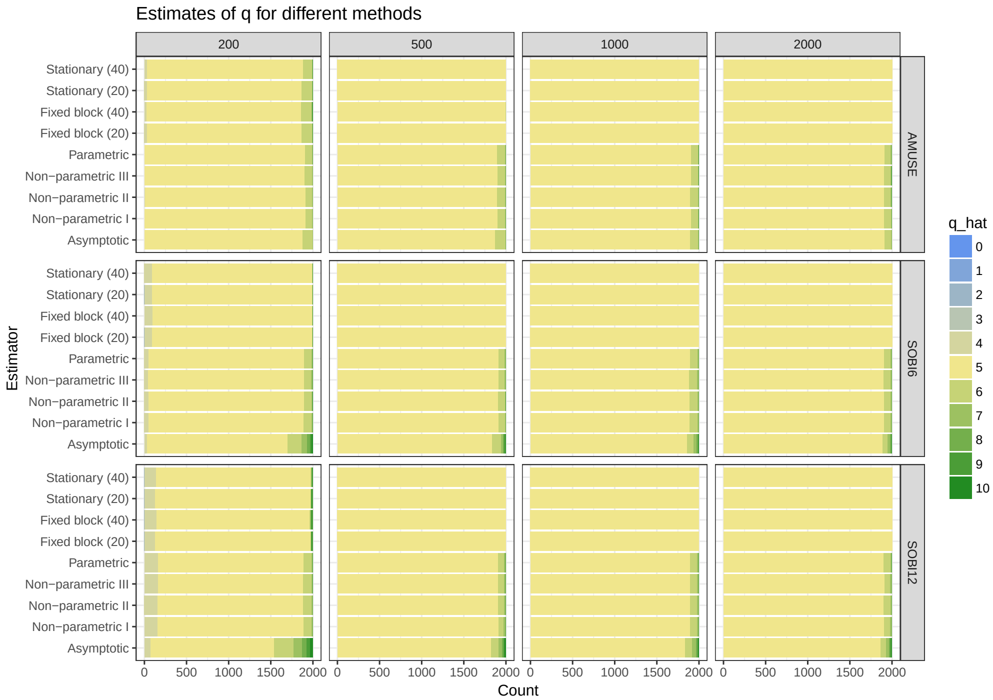

Figure S2: Estimating q by divide-and-conquer in Setting D1t.

Setting D3t: Five MA(2) processes with parameters (0.1, 0.1) having univariate t_5-distributed innovations together with five t_5-distributed white noise processes.

In all three cases the t_5-distributions were standardized to have unit variances.

Figure S3: Estimating q by divide-and-conquer in Setting D2t.

Figure S4: Estimating q by divide-and-conquer in Setting D3t.

Bibliography

Brockwell, P. and R. Davis (1991). *Time Series: Theory and Methods*. Springer.

Miettinen, J., K. Illner, K. Nordhausen, H. Oja, S. Taskinen, and F. Theis (2016). Separation of uncorrelated stationary time series using autocovariance matrices. *Journal of Time Series Analysis 37*, 337–354.

Serfling, R. J. (2009). *Approximation Theorems of Mathematical Statistics*, Volume 162. John Wiley & Sons.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Aalto University, Finland

E-mail: joni.virta@aalto.fi

Vienna University of Technology

E-mail: klaus.nordhausen@tuwien.ac.at